UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/621,474	07/17/2003	Hiroshi Nomiyama	JP920020109US1	2462

53792   7590   07/30/2007
DILLON & YUDELL LLP
8911 N. CAPITAL OF TEXAS HWY.
SUITE 2110
AUSTIN, TX 78759

EXAMINER
ROSE, HELENE ROBERTA

ART UNIT	PAPER NUMBER
2163	

MAIL DATE	DELIVERY MODE
07/30/2007	PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

PTOL-90A (Rev. 04/07)

	Application No.	Applicant(s)
Office Action Summary	10/621,474	NOMIYAMA ET AL.
	Examiner	Art Unit	
	Helene Rose	2163	

-- *The MAILING DATE of this communication appears on the cover sheet with the correspondence address* --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE *3* MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.
- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133).
  Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *5/25/2007*.

2a)☒ This action is **FINAL**.   2b)☐ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *15-17 and 21-24* is/are pending in the application.

    4a) Of the above claim(s) *1-14 and 18-20* is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *15-17 and 21-24* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☒ The drawing(s) filed on *17 July 2003* is/are: a)☒ accepted or b)☐ objected to by the Examiner.

    Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

    Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☒ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

    a)☒ All   b)☐ Some * c)☐ None of:

      1.☒ Certified copies of the priority documents have been received.

      2.☐ Certified copies of the priority documents have been received in Application No. _____.

      3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

    * See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1)☒ Notice of References Cited (PTO-892)

2)☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)

3)☒ Information Disclosure Statement(s) (PTO/SB/08)
Paper No(s)/Mail Date *5/26/05.*

4)☐ Interview Summary (PTO-413)
Paper No(s)/Mail Date. _____ .

5)☐ Notice of Informal Patent Application

6)☐ Other: _____ .

Detailed Action

1. In response to communication entered on 5/25/2007, Claims 1-14 and 18-

20 are cancelled. Claims 15, 17, 21, and 24 have been amended.

2. Applicant's arguments filed with respect to the rejected claims have been

fully considered but they are not persuasive.

Information Disclosure Statement

3. The information disclosure statement filed 5/26/05 fails to comply with 37

CFR 1.98(a)(2), in which the cited JP foreign patent documents 2001-060165;

2001-325272; 07-006076 as well as the non-patent literature documents cited

as other art; requires a legible copy of each cited foreign patent document; each

non-patent literature publication or that portion which caused it to be listed; and

all other information or that portion which caused it to be listed. Applicant is

required to indicate on the Information Disclosure Form what is to be considered

whether it's the abstract or full document as it relates to the foreign patents,

neither was specified nor cited on the form, and a full translation of the non-

patent literatures must be submitted. Thus, the IDS has been placed in the

application file, but the information referred to therein has not been considered.

Claim Rejections − 35 U.S.C 103

4.   The following is a quotation of 35 U.S.C. 103(a) which forms the basis for

all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically
> disclosed or described as set forth in section 102 of this title, if the
> differences between the subject matter sought to be patented and the
> prior art are such that the subject matter as a whole would have been
> obvious at the time the invention was made to a person having ordinary
> skill in the art to which said subject matter pertains. Patentability shall
> not be negatived by the manner in which the invention was made.

5.   Claims 15-17 and 21-24 are rejected under 35 U.S.C. 103(a) as being

unpatentable over Mantha et al. (US Patent No. 6,163,779, Date of Patent:

December 19, 2000, hereinafter Mantha) in view of Wyler (US Patent No.

7,047,033, Date Filed: January 31, 2001).

Claims 15 and 21:

Regarding claims 15 and 21, discloses a method/and a computer readable

medium encoded with a computer program utilizing the same functionalities,

wherein Mantha teaches a method/and a computer readable medium encoded

with a computer program, wherein the computer program, when executed,

performs the steps of:

reading an HTML document of a web page as an analyzing object (Figure

14, all features, wherein its further defined in column 12, lines 42-47, wherein

the upper portion of the code represents the original HTML source code,

wherein in lower portion each if the href tags has been modified to point to the

local storage, which is interpreted to be equivalent to " reading the source

codes of said web pages from said storage device' , as described in paragraph

[0018], of applicant specification, Figure 21, Mantha);

conducting a temporary block analysis based on a description of HTML

tags of the HTML document (column 2, lines 29-38,wherein the original page,

i.e. the base HTML document, is then parsed to prepare a list of hypertext

references, wherein this is interpreted to be " conducting a block analysis" ,

wherein such references are typically represented by <a href> markup tags,

wherein for each reference tag in the base HTML document, wherein this is

interpreted to be " description of HTML tags of the HTML document" , that is

an embedded object, e. g., an image, a copy of that file is retrieved from the

server and then saved on the local hard drive, wherein the new HTML page, the

path name to the stored file is substituted for the original hypertext reference,

wherein " substitute" is interpreted to be equivalent to " temporary" , and

therefore, interpreted to be equivalent to " conducting a temporary block

analysis based on a description of HTML tags of the HTML document" ,

Mantha);

using the HTML tags to temporarily divide the HTML document into

blocks (column 2, lines 28-30, wherein the original page, i.e., the base HTML

document, is then parsed to prepare a list of hypertext references, wherein the

term " parsed" is interpreted to be a method of " dividing" , and lines 37-38,

wherein the path name to the stored file is substituted for the original hypertext

references, which is interpreted to be equivalent to " using the HTML tags to

temporarily divide the HTML document into blocks" , Mantha);

Mantha does not teach: identifying unnecessary information elements in

the HTML document" , wherein the unnecessary information elements are plural

information elements that include:

plural information elements that include an OBJECT_IMAGE having a same

Uniform Resource Locator (URL), wherein the OBJECT_IMAGE describes a type

of media used to display the HTML document, a block of text in the HTML

document that is shorter than a maximum predetermined length, and wherein the

block of text appears in the HTML document more than a predetermined

frequency, multiple anchors having a same title, image tags and only perform a

role of punctuation for text in the HTML document, and multiple text blocks

having a same description;

Mantha does not teach: defining any block in the HTML document that is

deemed to be meaningless as an OBJECT_DELIMINATOR, wherein a block is

deemed to be meaningless if that block contains only said unnecessary

information elements and at least one anchor; and crawling only anchors found

in blocks that have not been defined as OBJECT_DELIMATORs.

Mantha does not teach: defining any block in the HTML document that is

deemed to be meaningless as an OBJECT_DELIMINATOR, wherein a block is

deemed to be meaningless if that block contains only said unnecessary

information elements and at least one anchor.

Nor does Mantha teach: crawling only anchors found in blocks that have

not been defined as OBJECT DELIMITER' s.

On the other hand, Wyler teaches " identifying unnecessary information

elements in the HTML document" (column 12, lines 4-8, wherein the application

removes irrelevant information, such as images and data, i.e., advertising

banners, links to unrelated issues, from the webpage, wherein a web page is a

document written in Hypertext Markup Language, which is interpreted to be

equivalent to an " HTML" document, Wyler), wherein the unnecessary

information elements are plural information elements that include:

plural information elements that include an OBJECT_IMAGE having a same

Uniform Resource Locator (URL), wherein the OBJECT_IMAGE describes a type

of media used to display the HTML document (column 22, lines 20-28, wherein

Banner Advertisements, in which advertisement objects that appear in the

document in the form of banners, which may include image and/or links, and

Image Advertisements wherein these images that appear in the HTML page with

no relevance to the page subject; and column 22, line 57, wherein this is a

background image of the web/HTML page, wherein background image is

interpreted to be the static image that appears behind text, graphics, and other

web page components, which is interpreted to be equivalent to " plural

information elements that include an OBJECT_IMAGE having a same Uniform

Resource Locator (URL), wherein the OBJECT_IMAGE describes a type of media

used to display the HTML document" , Wyler),

 a block of text in the HTML document that is shorter than a maximum

predetermined length (column 23, lines 20-25, wherein these links to additional

text segments that are considered relevant, but they do match the user retrieve

range, i.e., site depth, and etc, which is interpreted to correspond to "a block of

text in the HTML document that is shorter than a maximum predetermined

length and column 23, lines 49-51, wherein the user is able to set specific

filtering criteria for some objects in order to enhance the application sensitivity

to specific objects to include or exclude these objects column 32, lines 53-57,

wherein if the base object is not very big, e.g., falls below a threshold defining

the minimum size for a base object to generate a adequate size, Wyler) and

wherein the block of text appears in the HTML document more than a

predetermined frequency (column 15, lines 18-19, wherein converting the

webpage into objects involves dividing the webpage into regions wherein a

region can further be broken down into objects, wherein object is defined by

properties which involves occurrences (the number of alphanumeric strings

within a text object or table, which is interpreted to correspond to "wherein the

block of text appears in the HTML document more than a predetermined

frequency" and column 32, lines 47-52, wherein base object is selected which is

the largest object on the webpage, wherein " web page" is interpreted to be

equivalent to a HTML document, if there is a tie, i.e. if the largest two or more

objects are similar to a predetermined extent in size, " which is interpreted to

be equivalent to a " predetermined frequency" , then the object with the most

words, wherein this is equivalent to " text appears in the HTML document" , in

its typically deemed to the base object, in which is interpreted to be equivalent

to " wherein the text appears in the HTML document more than a predetermined

frequency, Wyler),

multiple anchors having a same title (column 16, lines 66-67, wherein the

data field in the image header contains the title that match the title of the base

object, wherein the term " match" is interpreted to correspond to the term

" same" , Wyler),

image tags that only perform a role of punctuation for text in the HTML

document (column 11, lines 43-61, wherein the application searches the web

page source or an input text file from markup languages, which is equivalent to a

" HTML document" , and wherein the application passes the page content to

one of the three following functions, which is interpreted to be equivalent to a

" role" , and wherein the markup language parses and analyze the markup

languages, and wherein the rich text format parses and analyze the text by

taking common knowledge of the text format, like bigger font size, which is

interpreted to be equivalent to " image tags that only perform a role of

punctuation for text in the HTML document" , Wyler), and

multiple text blocks having a same description (column 16, lines 1-4,

wherein image.words matching, and the image format contains a header with a

data field which describes the image content or the article that relates the

image, also refer to figure 2, wherein the object text identifies words matching,

Wyler);

Wyler teaches: defining any block in the HTML document that is deemed

to be meaningless as an OBJECT_DELIMINATOR, wherein a block is deemed to

be meaningless if that block contains only said unnecessary information

elements and at least one anchor (column 12, lines 4-8, wherein the application

removes irrelevant information from the webpage and reorganizes the

information into objects with categories in a file represent by the M2O script

language, which is interpreted to be equivalent to " defining any blocks in the

HTML document that is deemed to be meaningless as an OBJECT DELIMINTER,

and wherein irrelevant information includes links to unrelated issues, advertising

banners, and images, which is interpreted to be" wherein a block is deemed to

be meaningless if that block contains only unnecessary information elements and

at least one anchor" , Wyler);

Wyler teaches: crawling only anchors found in blocks that have not been

defined as OBJECT DELIMITER' s (column 14, lines 39-44, wherein application

searches all the documents for words that fit into the index category and when

finding such words the application program inserts an index command, in which

from that point on the web pages are called documents, which is interpreted to

be equivalent to " crawling only anchors found in blocks that have not been

defined as OBJECT DELIMITER' s" , Wyler).

It would have been obvious to one of the ordinary skill in the art at the

time of the invention to incorporate Wyler teachings into Mantha system. A

skilled artisan would have been motivated to combine as suggested by Wyler

[column 12, lines 1-2], in order to recognize irrelevant data.

As a result, establishing an improved method of providing tailored

information and a suitable display to a user.

Claims 16 and 22:

Regarding claims 16 and 22, most of the limitations have been noted in the

rejection of claims 15 and 21. In addition, the combination of Mantha and Wyler

teaches wherein the maximum predetermined length is 12 bytes (column 11,

lines 27-28, wherein the object which is the biggest or has the most number of

words in it; column 13, lines 58-61, wherein object size is a value equal to width

* height of the object; column 32, line 50, wherein predetermined " extent" is

interpreted to be equivalent to " length" , Wyler).

Claims 17 and 23:

Regarding claims 17 and 23, most of the limitations have been noted in the

rejection of claims 15 and 21. In addition, the combination of Mantha and Wyler

teaches wherein the predetermined frequency is ten times (column 16, lines 39-

41, wherein the mechanism for selecting the relevant objects is based on

selecting the objects with weights that pass the predefined thresholds, which is

interpreted to be equivalent to " predetermined frequency" ; and column 19,

lines 3-7 and 33-36, wherein the application tries to reduce the index list length

by finding identical words with different page numbers and wherein if chapters

and sections with titles and subtitles do not appear in the same document after

the third level only the following changes takes place, which is irrelevant

images/data are taken off; column 31, lines 50, wherein system counts the

number of words in the object which do not occur in the base object, wherein

the proportion of words in the object which occur in the base object from among

the total number of words in the object, which determines the word matching,

which is equivalent to frequency, Wyler).

<u>Claim 24</u>:

Regarding Claim 24, <u>SEE</u> claims 15 and 21 above, wherein this limitation is

substantially the same/or similar and therefore rejected under the same

rationale).

<div align="center"><u>Examiner responses to Applicant Arguments</u></div>

<u>Applicant States</u>:

This Amendment is submitted in response to the Office Action dated February

22, 2007, 2006, having a shortened statutory period set to expire May 22, 2007.

The present amendment amends Claims 15, 21 and 24. Upon entry of the

proposed claims, Claims 15-17 and 21-24 will now be pending.

<u>Rejections 35 U.S.C. § 103</u>

In paragraph 4 of the present Office Action, Claims 15-17 and 21-24 are

rejected as being unpatentable over Mantha et al. (U.S. Patent No. 6,163,779 -

"Mantha") in view of Wyler (U.S. Patent No. 7,047,033 - "Wyler"). Applicants

respectfully traverse these rejections.

With regards to exemplary Claim 15, Applicant Argues (1): <u>a combination</u>

<u>of the cited art does not teach or suggest "identifying unnecessary information</u>

<u>elements in the HTML document, wherein the unnecessary information elements</u>

<u>include...a block of text in the HTML document that is shorter than a maximum</u>

<u>predetermined length, and wherein the block of text appears in the HTML</u>

<u>document more than a predetermined frequency</u>," as supported at paragraph

[0085] of the current disclosure. That is, unnecessary information includes

multiple repetitions of a same short block of text. The Examiner cites Wyler on

col. 31, line 65 to col. 32, line 7, and col. 32, lines 47-52, which state:

The "logical location" of an object which is interiorly disposed relative to

the base object is the maximum value e.g. 100. The "logical location" of any

other object is the distance, on the webpage, of that object from the base

object."

"Classifying one or more objects as cardinal: As described, a base object

is selected which is the largest object on the webpage. If there is a tie, i.e. if the

largest two or more objects are similar, to a predetermined extent, in size, then

the object with the most words in it is typically deemed to be the base object."

Thus, these passages teach that the location of an object can be described

as a distance to the largest object on a webpage. Applicants respectfully

traverse the Examiner's position that this is equivalent to multiple short blocks

of text being located on an HTML document too many times.

Examiner Response to Applicant Argument (1):

Examiner is not persuaded. Referring See Wyler, wherein the prior art of

record does teach identifying unnecessary information elements in the HTML

document ~ column 12, lines 4-8, wherein the application removes irrelevant

information, such as images and data, i.e., advertising banners, links to unrelated

issues, from the webpage, wherein a web page is a document written in

Hypertext Markup Language, which is interpreted to be equivalent to an

" HTML" document, wherein the unnecessary information elements include...a

block of text in the HTML document that is shorter than a maximum

predetermined length ~ column 32, lines 53-57, wherein if the base object is not

very big, e.g., falls below a threshold defining the minimum size for a base

object to generate a adequate size, and wherein the block of text appears in the

HTML document more than a predetermined frequency ~ column 15, lines 18-19,

wherein converting the webpage into objects involves dividing the webpage into

regions wherein a region can further be broken down into objects, wherein

object is defined by properties which involves occurrences (the number of

alphanumeric strings within a text object or table, which is interpreted to

correspond to "wherein the block of text appears in the HTML document more

than a predetermined frequency", and also see column 32, lines 47-52, wherein

base object is selected which is the largest object on the webpage, wherein

" web page" is interpreted to be equivalent to a HTML document, if there is a

tie, i.e. if the largest two or more objects are similar to a predetermined extent

in size, " which is interpreted to be equivalent to a " predetermined

frequency" , then the object with the most words, wherein this is equivalent to

" text appears in the HTML document" , in its typically deemed to the base

object, in which is interpreted to be equivalent to " wherein the text appears in

the HTML document more than a predetermined frequency.

Furthermore, a combination of the Applicant Argues (2) that the <u>cited art</u>

<u>does not teach or suggest "a block is deemed to be meaningless if that block</u>

<u>contains only said unnecessary information elements and at least one anchor," as</u>

<u>supported in paragraph [0085] of the present specification.</u>

The Examiner cites Wyler on col. 12, lines 4-8, which states:

In this level the application removes irrelevant information (images and

data i.e. advertising banners, links to unrelated issues) from the webpage, and

reorganizes the information into objects with categories in a file represent by

the M20 script language.

That is, Wyler teaches that images such as advertising banners and links

to other irrelevant pages can be purged from a webpage. However, there is no

teaching or suggestion of "multiple short block of text being repeated" (as

discussed above), "multiple anchors having a same title," "image tags that only

perform a role of punctuation," and "text block having a same description" as

being requisite components for defining a meaningless block in an HTML

document.

Examiner Response to Applicant Argument (2):

Examiner is not persuaded. See Wyler ~ column 12, lines 4-8, wherein

the application removes irrelevant information such as images and data, i.e.,

advertising banners, links to unrelated issues from the webpage, wherein links is

interpreted to correspond to " anchors" , and reorganizes the information into

objects with categories in a file represent by the M2O script language, which is

interpreted to be equivalent to " defining any blocks in the HTML document that

is deemed to be meaningless as an OBJECT DELIMINTER, and wherein

irrelevant information includes links to unrelated issues, advertising banners,

and images, which is interpreted to be" wherein a block is deemed to be

meaningless if that block contains only unnecessary information elements and at

least one anchor" .

Similarly, Applicant Argues (3) there is no teaching or suggestion that this

meaningless block in the HTML document has "at least one anchor."

Examiner Response to Applicant Argument (3):

Examiner is not persuaded. See Wyler ~ column 12, lines 4-8, wherein

the application removes irrelevant information (images and data, i.e., advertising

banners, links to unrelated issues from the webpage), wherein "links" is

interpreted to be equivalent to "anchors".

Furthermore, Applicant Argues (4) a combination of the cited art does not

teach or suggest, "crawling only anchors found in blocks that have not been

defined as OBJECT DELIMETERS" (i.e., only crawling anchors that are not

meaningless). The Examiner cites col. 14, lines 39-44 of Wyler for this teaching.

This passage states: Second, the application searches all the documents for

words that fit into the Index category, and when finding such words, the

application inserts an "index" command. From that point on, the web pages are

called "documents" since they have no longer have properties of a webpage.

Examiner Response to Applicant Argument (4):

Examiner is not persuaded. See Wyler ~ column 14, lines 15-25, wherein

the M20 script begins with scanning the entire webpage and parsing the contents

into words related to the webpage commands and words relating to the user

relevant information and some of the commands that are found may be relevant

for formatting a document/webpage in a book-style format/webpage for devices

with screen size and browser limitations and some may be irrelevant (e.g.

remarks, search engine keywords, etc), wherein the relevant commands that are

found are translated into M20 Script language and column 14, lines 39-44,

wherein application searches all the documents for words that fit into the index

category and when finding such words the application program inserts an index

command, in which from that point on the web pages are called documents,

which is interpreted to be equivalent to " crawling only anchors found in blocks

that have not been defined as OBJECT DELIMITER' s" .

This passage states that a crawler ("application") searches all documents

for key words ("that fit into the Index category"). Applicant Argues (5) There is

no teaching or suggestion of crawling only anchors that are not composed

exclusively of "plural information elements that include an OBJECT IMAGE having

a same Uniform Resource Locator (URL), wherein the OBJECT IMAGE describes

a type of media used to display the HTML document, a block of text in the HTML

document that is shorter than a maximum predetermined length, and wherein the

block of text appears in the HTML document more than a predetermined

frequency, multiple anchors having a same title, image tags that only perform a

role of punctuation for text in the HTML document, and multiple text blocks

having a same description."

Examiner Response to Applicant Argument (5):

In addition some of applicants are arguments are substantially the same as

applicant arguments (1). Again, Examiner is not persuaded. See Wyler, wherein

the prior art of record teaches "plural information elements that include an

OBJECT IMAGE having a same Uniform Resource Locator (URL), wherein the

OBJECT IMAGE describes a type of media used to display the HTML document ~

column 22, lines 20-28, wherein Banner Advertisements, in which advertisement

objects that appear in the document in the form of banners, which may include

image and/or links, and Image Advertisements wherein these images that appear

in the HTML page with no relevance to the page subject; and column 22, line 57,

wherein this is a background image of the web/HTML page, wherein background

image is interpreted to be the static image that appears behind text, graphics,

and other web page components, which is interpreted to be equivalent to

" plural information elements that include an OBJECT_IMAGE having a same

Uniform Resource Locator (URL), wherein the OBJECT_IMAGE describes a type

of media used to display the HTML document" , a block of text in the HTML

document that is shorter than a maximum predetermined length ~ column 23,

lines 20-25, wherein these links to additional text segments that are considered

relevant ,but they do match the user retrieve range, i.e., site depth, and etc,

which is interpreted to correspond to "a block of text in the HTML document

that is shorter than a maximum predetermined length and column 23, lines 49-

51, wherein the user is able to set specific filtering criteria for some objects in

order to enhance the application sensitivity to specific objects to include or

exclude these objects and column 32, lines 53-57, wherein if the base object is

not very big, e.g., falls below a threshold defining the minimum size for a base

object to generate a adequate size, and wherein the block of text appears in the

HTML document more than a predetermined frequency ~ See column 16, lines

39-41, wherein the mechanism for selecting the relevant objects on selecting

the objects with weights that pass the predefined thresholds which is interpreted

to correspond to "predetermined frequency", column 23, lines 59-60, wherein

the keywords are selected in accordance to occurrence and significance (words

that appear in titles, bold, etc) and column 32, lines 47-52, wherein base object

is selected which is the largest object on the webpage, wherein " web page" is

interpreted to be equivalent to a HTML document, if there is a tie, i.e. if the

largest two or more objects are similar to a predetermined extent in size,

" which is interpreted to be equivalent to a " predetermined frequency" , then

the object with the most words, wherein this is equivalent to " text appears in

the HTML document" , in its typically deemed to the base object, in which is

interpreted to be equivalent to " wherein the text appears in the HTML

document more than a predetermined frequency, multiple anchors having a same

<u>title, image tags that only perform a role of punctuation for text in the HTML</u>

<u>document, and multiple text blocks having a same </u>description ~ see column 16,

lines 66-67, wherein the data field in the image header contains the title that

match the title of the base object, wherein the term " match" is interpreted to

correspond to the term " same" , and column 16, lines 1-4, wherein

image.words matching, and the image format contains a header with a data field

which describes the image content or the article that relates the image, also

refer to figure 2, wherein the object text identifies words matching

<u>Applicant States:</u>

Therefore, in light of the present amendment further distinguishing the

definition of "unnecessary information," Applicants respectfully request that the

rejection of Claims 15, 21 and 24 be withdrawn. With regards to exemplary

Claim 16, a combination of the cited art does not teach or suggest, "the

maximum predetermined length (of the block of text) is 12 bytes." For this

feature, the Examiner cites Wyler at col. 11, lines 27-28; col. 13, lines 58-61;

and col. 32, line 50, which state:

Base - The object, which is the biggest or has the most Object number of

words in it.

Object size - the object size is a value equal to Width * Height of the

object

(T)he object with the most words in it is typically deemed to be the base

object

The cited passages state that the biggest size of a base object (see above

for discussion of what a "base object" denotes) may be determined. Applicants

respectfully traverse the Examiner's statement that this is equivalent to a block

of text having a maximum size (12 bytes), which is used to denote the block of

text as being "unnecessary."

Applicants therefore request that the rejection of Claims 16 and 22 be

withdrawn.

Similarly, with regards to exemplary Claim 17, Applicant Argues (6) <u>a</u>

<u>combination of the cited art does not teach or suggest, "the predetermined</u>

<u>frequency (of occurrences of the block of text) is ten times</u>." For this feature, the

Examiner cites Wyler on col. 16, lines 39-41; col. 19, lines 3-7 and 33-36; and

col. 31, line 50, which state:

<u>Examiner Response to Applicant Argument (6):</u>

Examiner is not persuaded. Applicant is reminded that the claim language

presently defined within the application under review is interpreted to be

intended use and can be read/interpreted as data or any functionality that is

equivalent to the functionality being claimed. See Wyler ~ column 16, lines 39-

41, wherein the mechanism for selecting the relevant objects on selecting the

objects with weights that pass the predefined thresholds which is interpreted to correspond to "predetermined frequency" and column 23, lines 59-60, wherein the keywords are selected in accordance to occurrence and significance (words that appear in titles, bold, etc).

Applicant States:

The mechanism of selecting the relevant objects is based on selecting the objects with weights that pass the predefined thresholds. In FIG. 4 we can see (marked by diagonal lines) a relevant region that passes the predefined thresholds.

In this phase, the application tries to reduce the Index list length by finding identical words with different page numbers. The application then indicates the word followed by a list of all the reference page numbers. If Chapters and Sections with Titles and Sub-Titles do not appear in the document after the third level, only the following changes typically take place: 1. In the second level--irrelevant images/data are taken off.

Typically, the "word matching" property is computed by performing a key word matching process. In this process, each word within the object whose "word matching" property is being computed is taken up in turn and the system determines whether this word occurs in the base object. The system counts the number of words in the object, which do occur in the base object. The

proportion of words in the object, which occur in the base object, from among the total number of words in the object, typically determines the "word matching" property of the object.

Applicants understand the Examiner's position to be, in essence, that a passage can be determined as being significant to a crawler if a particular word is found multiple times in a base (large) object, and that this operation is equivalent to a block of text occurring more than ten times causing a block of an HTML document to be deemed meaningless. Applicants respectfully disagree, since the two concepts are diametrically opposed. That is, Wyler teaches that any passage having multiple entries (of a word on a document) is meaningful (as per any standard crawling technique). Conversely, the present invention states that multiple entries (of a block of data in a document) make that block of data meaningless.

Applicants therefore request that the rejection of Claims 17 and 23 be withdrawn.

Response to Arguments

Applicant's arguments filed on 5/25/2007, with respect to the rejected

claims in view of the cited references have been considered but are moot in

view of applicant's amended claims necessitate new ground(s) of rejection.

Prior Art of Record

1. Wyler et al   (US Patent No. 7,047,033)

2. Ishikawa et al   (US Patent No. 5,848,407)

3. Finseth et al   (US Patent No. 6,271,840)

4. Mantha et al   (US Patent No. 6,163,779)

5. Wang Baldonado   (US Patent No. 6,704,722)

Conclusion

Applicant's amendment necessitated the new ground(s) of rejection

presented in this Office action. Accordingly, THIS ACTION IS MADE FINAL.

See MPEP § 706.07(a). Applicant is reminded of the extension of time policy as

set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire

THREE MONTHS from the mailing date of this action. In the event a first reply

is filed within TWO MONTHS of the mailing date of this final action and the

advisory action is not mailed until after the end of the THREE-MONTH

shortened statutory period, then the shortened statutory period will expire on

the date the advisory action is mailed, and any extension fee pursuant to 37 CFR

1.136(a) will be calculated from the mailing date of the advisory action. In no

event, however, will the statutory period for reply expire later than SIX

MONTHS from the date of this final action.
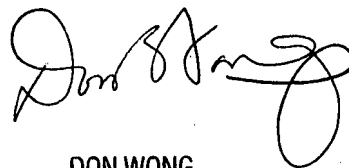

Point of Contact

Any inquiry concerning this communication or earlier communications

from the examiner should be directed to Helene Rose whose telephone number

is (571) 272-0749. The examiner can normally be reached on 8:00am - 4:30pm

Monday-Friday.

If attempts to reach the examiner by telephone are unsuccessful, the

examiner's supervisor, Don Wong can be reached on (571) 272-1834. The fax

phone number for the organization where this application or proceeding is

assigned is 571-273-8300.

Information regarding the status of an application may be obtained from

the Patent Application Information Retrieval (PAIR) system. Status information

for published applications may be obtained from either Private PAIR or Public

PAIR. Status information for unpublished applications is available through

Private PAIR only. For more information about the PAIR system, see http://pair-

direct.uspto.gov. Should you have questions on access to the Private PAIR

system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-

free). If you would like assistance from a USPTO Customer Service

Representative or access to the automated information system, call 800-786-

9199 (IN USA OR CANADA) or 571-272-1000.

HRR
Technology Center 2100
July 20, 2007   DON WONG
                                        SUPERVISORY PATENT EXAMINER
                                         TECHNOLOGY CENTER 2100